A Preliminary Examination of WRF Ensemble Prediction of Convective Mode Evolution

BRADLEY R. CARLBERG, WILLIAM A. GALLUS JR., AND KRISTIE J. FRANZ

Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

(Manuscript received 29 September 2017, in final form 30 March 2018)

ABSTRACT

Accurately simulating convective mode evolution can assist forecasters in the severe weather warning process. A few prior studies have examined the skill of simulating convective modes using single, deterministic forecasts. The present study extends the earlier evaluations to a small, four-member ensemble, with each member incorporating varying initial and lateral boundary conditions, microphysics schemes, and planetary boundary layer schemes. Simulated convective modes from thirty-two 12-h simulations were categorized into nine classifications using a classification scheme developed from previous studies. Multiple methods were used to derive forecasts of these convective classifications, creating an hourly deterministic ensemble mode forecast and probabilistic forecasts for 1-, 6-, and 12-h periods. Forecasts were compared with observed radar reflectivity for verification. In general, hourly deterministic ensemble mode forecasts showed improvement over individual member forecasts. The small ensemble produced more skillful individual cellular convective mode forecasts than individual linear mode forecasts, with the least skill present for bow echoes and squall lines with trailing stratiform precipitation. In contrast, the ensemble was more skillful at forecasting the broader linear convective group than the broader cellular convective group. For a limited number of these cases, a test was performed using a larger 10-member ensemble run by the National Center for Atmospheric Research (NCAR) to examine what impacts the small ensemble size might have. The results did not differ substantially, suggesting the findings from the small ensemble can be generalized. Probabilistic forecasts for longer time periods were more skillful than shorter-term probabilistic forecasts.

1. Introduction

Accurate simulation of convective mode evolution can provide additional useful guidance for severe weather forecasts and warnings because individual convective modes have been found to be associated with different severe storm report types. Tornadoes and hail of all sizes, for example, were the two main threats associated with cellular convection in Gallus et al. (2008, hereafter G08) and Duda and Gallus (2010). Of cellular convective mode classifications, discrete cells within a broken line were associated with the most tornadoes in two previous studies (G08; Smith et al. 2012). Damaging winds have been found to be the greatest threat from linear convection (G08), particularly with bow echoes (Fujita 1978). Linear systems with any associated stratiform precipitation (leading, parallel, or trailing) and broken-line cellular systems have been shown to be the greatest flash flood threats (Pettet and Johnson 2003; G08). The Pettet and Johnson (2003) and G08 conclusions are consistent with the Doswell et al. (1996) findings that slow-moving, training convection often causes flash floods.

Although a consensus exists that different convective modes favor different types of weather hazards, numerous different methods of classification have been proposed. Bluestein and Jain (1985), for example, used squall-line development from broken-line, back-building, broken-areal, and embedded-areal convection to classify convective systems. Parker and Johnson (2000) proposed trailing, parallel, and leading squall-line classifications based on stratiform precipitation location. Later, Fowle and Roebber (2003) added multicellular and isolated convective mode classifications to examine mode depiction in numerical models, and Done et al. (2004) applied a two-category classification scheme that separated convection into either quasi-linear or non-squall-line categories. An automated classification process developed by Baldwin et al. (2005) categorized events as nonconvective, convective linear, and convective cellular based on observed hourly precipitation amounts. Baldwin et al. (2005) further

Corresponding author: Bradley R. Carlberg, bcarl@iastate.edu

DOI: 10.1175/WAF-D-17-0149.1

separated cellular convection into isolated cell and multicell cluster categories, and G08 identified clusters of cells as a separate category from isolated cells or lines of cells. When convection was neither linear nor cellular, Grams et al. (2006) categorized it as continuous nonlinear. G08 and Duda and Gallus (2010) used many of these convective modes and added the classification of a squall line with no stratiform precipitation while relating severe storm reports to each convective mode. Because G08 used a classification scheme that provides a comprehensive number of categories built from the previous studies (five linear, three cellular, and one nonlinear), their scheme was used to classify convective systems in the present study.

Convection-allowing model simulations are now commonly run and have the ability to resolve finer-scale convective system details that are similar to those seen with observational radar (Done et al. 2004; Kain et al. 2006). However, Done et al. (2004) and Snively and Gallus (2014, hereafter SG14) have shown that the Weather Research and Forecasting (WRF) Model still struggles with producing some linear convective modes. Done et al. (2004) showed the WRF Model did not simulate enough stratiform rain regions, while SG14 found that the WRF Model poorly simulated squall lines with trailing stratiform precipitation and bow echoes. These previous studies focused on convective mode prediction using single deterministic forecasts. To address the uncertainty in a given forecast, combinations of model configurations are being increasingly used to create ensemble forecasts (Toth and Kalnay 1993; Elmore et al. 2003; Homar et al. 2006; Leutbecher and Palmer 2008; Clark et al. 2009). Historically, ensemble mean forecasts have typically performed better than individual members (Leith 1974; Fritsch et al. 2000; Baars and Mass 2005), and ensembles facilitate probabilistic forecasts (Murphy 1991). However, the use of an ensemble to predict the convective mode poses some additional challenges as the events being predicted fall into a set number of categories that cannot be numerically averaged. An example of a similar challenge can be found in Wandishin et al. (2005), who created postprocessing algorithms to produce precipitation-type forecast probabilities.

The ability for a small, four-member WRF ensemble to improve short-term convective mode forecasts is the focus of the present study. Although such a small convection-allowing ensemble has potential for several shortcomings, the present study can be thought of as a test to provide an initial look into the process of applying an ensemble forecast to convective modes. Twelve-hour simulations with a convection-allowing 4-km horizontal grid were completed to produce forecasts using the WRF Model, version 3.6.1, with the Advanced Research dynamics core (ARW; Skamarock et al. 2008). The model was chosen because it could be used in a real-time setting, which was necessary to support a project goal of providing information to a local National Weather Service office for consideration during operational forecasting situations. To produce forecasts for real-time situations, the WRF simulations were initialized while convection was occurring, or about to occur, in the U.S. Upper Midwest study domain (Fig. 1). The study period covered the 2015 and 2016 warm seasons (March–September). A total of 32 simulations covering 21 days (Table 1) were investigated based on convection occurring continuously within the domain for a minimum of 6 h, matching the persistence requirement of Done et al. (2004). Because of the potential shortcomings with a small ensemble, a limited analysis of a larger ensemble was also conducted to test how sensitive convective mode forecasts are to ensemble size. Details of the ensemble setup, convective mode classification, forecast creation, and forecast verification methods are explained in section 2. Results from the forecasts are described in section 3. Finally, conclusions are presented in section 4.

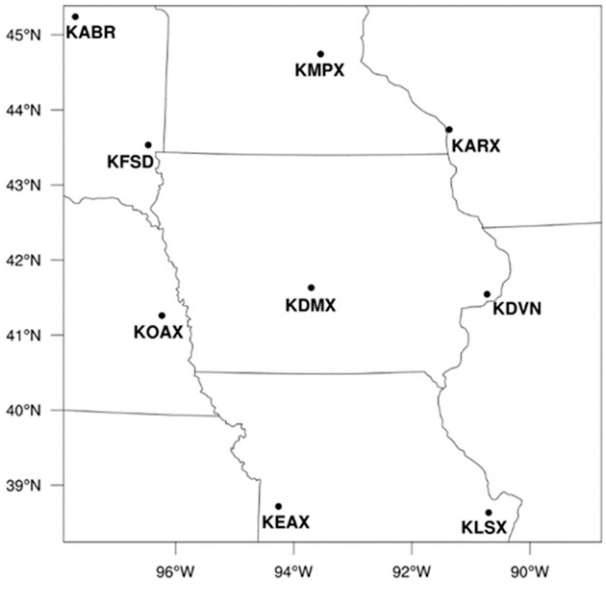

FIG. 1. The domain for which ensemble convective mode forecasts were tested including the nine radar locations used for data assimilation.

2. Methodology

a. Ensemble generation and classification

A small WRF ensemble that varies the sources of initial and lateral boundary conditions was used to create convective mode forecasts (Table 2). The ensemble was restricted to four members because of the limited

TABLE 1. Initialization times and dates for all 32 simulations.

2015	
1200 UTC 9 Apr	1800 UTC 28 Jul
2100 UTC 18 Apr	0000 UTC 29 Jul[a]
0000 UTC 4 May[a]	0600 UTC 9 Aug
0000 UTC 5 May[a]	0000 UTC 10 Aug[a]
1200 UTC 7 May	0000,[a] 0300, and 0600 UTC 17 Aug
1200 UTC 24 May	0300, 0600, 0900, and 1200 UTC 18 Aug
0300 UTC 7 Jun	0300, 0600, and 0900 UTC 28 Aug
0600, 1200, and 1800 UTC 24 Jun	2016
0000 UTC 25 Jun[a]	2100 UTC 23 Mar
1200 UTC 11 Jul	0000 UTC 6 Apr[a]
0600 and 1200 UTC 16 Jul	1200 UTC 9 May

[a] Case used for the NCAR comparison.

computational resources available. Because the intent of these forecasts was to provide real-time information to National Weather Service forecasters, it was important that the ensembles could be generated in a short time frame. Therefore, 3- or 6-h model forecast output from the NAM and GFS models was used for the initial conditions, and subsequent forecasts were used for lateral boundary conditions. For example, a simulation initialized at 0600 UTC used the 6-h forecast output from the 0000 UTC NAM and GFS runs for initial conditions with the 9-, 12-, 15-, and 18-h forecasts used for lateral boundary conditions. Because the number of ensemble members used for an initial look into applying an ensemble forecast to convective modes was limited to four, the members were designed to provide as diverse a set of forecasts as possible without using more sophisticated methods of perturbing the initial and lateral boundary conditions. Two of the members used the same model forecast output for both initial and lateral boundary conditions; the other two members mixed the model forecast output with initial conditions supplied from one model and lateral boundary conditions from the other model (hereafter the individual members will be referred to using the syntax of initial conditions/lateral boundary conditions: NAM/NAM, GFS/GFS, NAM/GFS, and GFS/NAM). Each of the four configurations also featured different microphysics and planetary boundary layer scheme combinations (Table 2). The NAM/NAM member was paired with the Thompson two-moment microphysics scheme (Thompson et al. 2008) and the Yonsei University (YSU) planetary boundary layer (PBL) scheme (Hong et al. 2006). The GFS/GFS member combined the Morrison two-moment microphysics scheme (Morrison et al. 2009) and the Mellor–Yamada–Nakanishi–Niino level 2.5 (MYNN) PBL scheme (Nakanishi and Niino 2009). The NAM/GFS member used the Goddard microphysics scheme (Tao et al. 1989) with the Quasi-Normal Scale Elimination (QNSE) PBL scheme (Sukoriansky et al. 2005). Finally, the GFS/NAM member combined the WRF single-moment 6-class (WSM6) microphysics scheme (Hong and Lim 2006) and the Mellor–Yamada–Janjić (MYJ) PBL scheme (Janjić 1994).

To reduce possible model spinup problems in the first few hours of a forecast, raw NEXRAD level II radar data were assimilated with 4-km resolution at the initial time step of the forecast for all four members. Radar data assimilation has also been shown to improve quantitative precipitation forecasts (QPFs) during the first 6–8 h of a forecast (Berenguer et al. 2012; Stratman et al. 2013). The radar data from nine WSR-88D radars covering the 800 km × 800 km domain centered on Des Moines, Iowa, in the Upper Midwest were used (Fig. 1). These nine radars include Aberdeen, South Dakota (KABR); Sioux Falls, South Dakota (KFSD); Omaha, Nebraska (KOAX); Kansas City, Missouri (KEAX); Saint Louis, Missouri (KLSX); Lacrosse, Wisconsin (KARX); Twin Cities, Minnesota (KMPX); Davenport, Iowa (KDVN); and Des Moines, Iowa (KDMX). The radar data were assimilated by the Advanced Regional Prediction System (ARPS) three-dimensional variational data assimilation (3DVAR) program and the ARPS Data Analysis System (ADAS; Brewster 1996). Hydrometeors and cloud fields were adjusted based off the radar reflectivity data through the use

TABLE 2. WRF Model configurations of each ensemble member from the present study and SG14.

	Member 1	Member 2	Member 3	Member 4	SG14
Initial conditions	NAM	GFS	NAM	GFS	NAM
Lateral boundary conditions	NAM	GFS	GFS	NAM	NAM
Microphysics scheme	Thompson	Morrison	Goddard	WSM6	Thompson
PBL scheme	YSU	MYNN 2.5	QNSE	MYJ	MYJ
Longwave radiation scheme	Goddard	Goddard	RRTMG	Goddard	RRTM
Shortwave radiation scheme	Goddard	Goddard	RRTMG	Goddard	Dudhia
Land surface scheme	Noah	Noah	Noah	RUC	Noah
Surface-layer scheme	MM5	MM5	QNSE	Eta	Eta

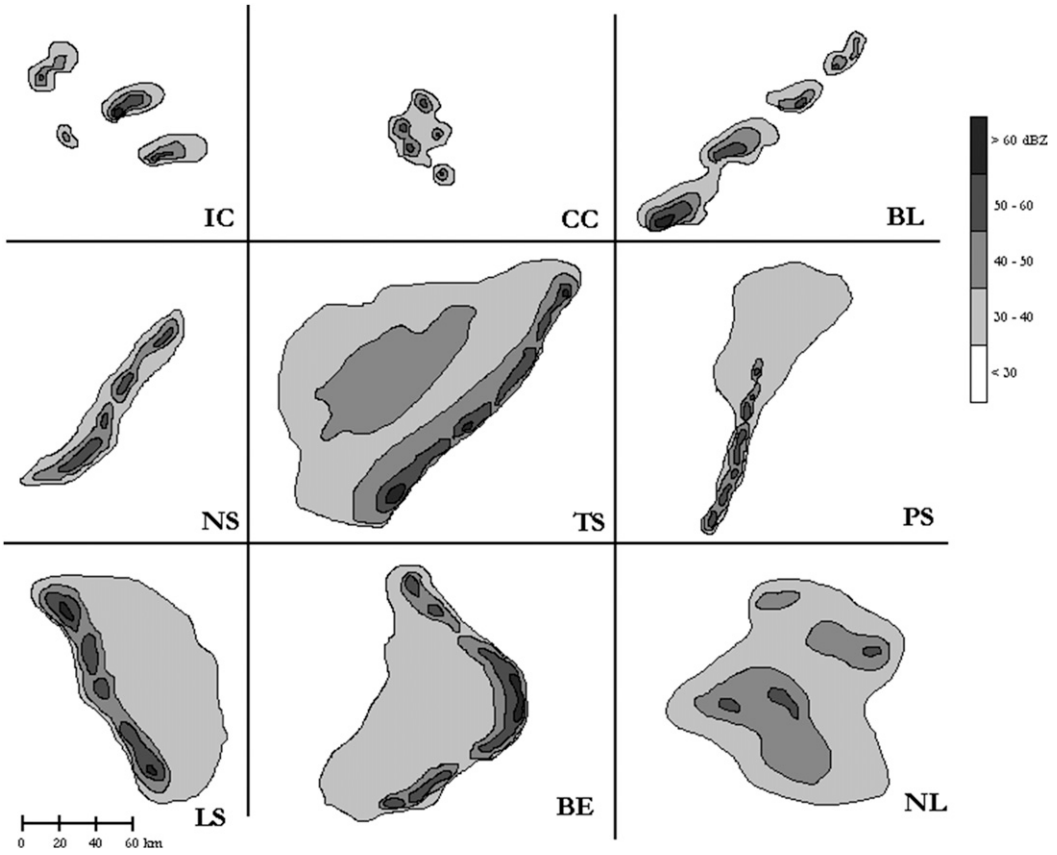

FIG. 2. Schematic from G08 showing the nine convective modes used for classification.

of a cloud analysis procedure, a component of both ADA and ARPS 3DVAR, and the radial velocity data were analyzed by the three-dimensional variational scheme (Hu et al. 2006; Moser et al. 2015; Yan and Gallus 2016). Furthermore, external data are interpolated onto the ARPS grid, and the ARPS program reduces the initial oscillations in the pressure field by adjusting the wind fields to ensure the anelastic mass continuity equation is satisfied (CAPS 2017). Reducing the initial oscillations in the pressure field helps to mitigate any model spinup issues that might arise with mixing NAM and GFS initial and lateral boundary conditions.

Simulated radar reflectivity was output hourly from each model run and used to classify the convective system(s) present based on the nine classifications used in G08 (Fig. 2): isolated cells (IC), cluster of cells (CC), broken line (BL), squall line with no stratiform precipitation (NS), squall line with parallel stratiform precipitation (PS), squall line with leading stratiform precipitation (LS), squall line with trailing stratiform precipitation (TS), bow echoes (BE), and nonlinear systems (NL). Each of the nine classifications was also grouped into three broader categories: cellular convection consisting of IC, CC and BL events; linear convection including NS, PS, LS, TS, and BE events; and NL. Observed systems were classified in the same manner using radar reflectivity from the University Corporation for Atmospheric Research's (UCAR) image archive (information can be found at http://www2.mmm.ucar.edu/imagearchive/). While there is inherently some subjectivity built into the classification, guidelines used in G08, Duda and Gallus (2010), and SG14 were followed closely. The threshold used for convection was 40 dBZ over an area of at least 8 km $\times$ 8 km, and the threshold for stratiform precipitation was 30 dBZ (Hilgendorf and Johnson 1998). In order for a system to be considered linear, the convection needed to be at least 75 km long and have a 3:1 length-to-width ratio (G08). In addition, stratiform precipitation associated with linear convection needed to be at least twice as wide as the convective precipitation to be classified as TS, LS, or PS. Furthermore, as in SG14, characteristics of a convective mode needed to occur for at least two consecutive hours in order to receive that classification. Because of this consecutive hour requirement, a small fraction (13%) of the hourly forecasts are not truly independent samples.

A four-member ensemble would likely be smaller than those used operationally. Therefore, to get better insight into how the results from this study apply to other ensembles, the more sophisticated, 10-member National Center for Atmospheric Research (NCAR) ensemble (Schwartz et al. 2015) was also evaluated. The NCAR ensemble forecasts were initialized once daily at 0000 UTC (information can be found at http://ensemble.ucar.edu/about.php), matching the initialization times from seven of the small WRF ensemble events. Thus, these seven events were chosen for evaluation (Table 1).

b. Forecast approaches

Multiple methods were used to create convective mode forecasts based on the ensemble output. The first method created a deterministic forecast by selecting the statistical mode from the four members for each forecast hour, because the categorical convective modes cannot be numerically averaged in a way to produce a traditional ensemble mean forecast. For example, if three ensemble members produced a CC classification and the fourth produced a BL classification at a given hour, the ensemble mode was assigned as CC. Once the ensemble modes were determined, the forecast method followed the procedure in SG14 with a few modifications. The simulated convective modes immediately before and after the hour of interest were used to determine the mode when a tie occurred among the four ensemble members. When the same ensemble mode occurred in the prior and post hours, the convective mode at the time of the tie was classified to match the mode from the surrounding hours. If the ensemble mode matched just one of the prior or post hours, the mean was reclassified to match that convective mode. Because of the tendency for TS and BE to be underpredicted (Done et al. 2004; SG14), two ensemble mode bias corrections were used to favor TS and BE with the small WRF ensemble. First, when at least two of the four members were classified as TS (BE), the ensemble mode was set to TS (BE). Second, the ensemble mode was set to TS (BE) when that particular hour forecast at least one member producing a TS (BE). After the bias corrections were made, the requirement of a mode needing to persist for two consecutive hours was still enforced.

The second forecast method produced hourly probabilistic convective mode forecasts using two different techniques. In one, forecast probabilities of a convective mode were calculated from convective modes simulated for that particular hour, hereafter referred to as the direct forecast method. Because Duda and Gallus (2013) and SG14 found that timing errors were often present in convective mode forecasts, a 3-h neighborhood forecast method was also used to produce probabilistic forecasts for a given hour, hereafter referred to as the neighborhood forecast method. The neighborhood forecast method calculated probabilities of convective mode using the ensemble members for the current hour, 1 h prior, and 1 h post. For consistency, the neighborhood forecast method was only used when output from all three hours was available, thus, the first and last hours of a simulation were excluded.

The third forecast method created probabilistic forecasts of all nine convective modes occurring during a 6- or 12-h period. Probabilities of each mode were calculated for the entire 12-h forecast, the first 6 h of the forecast, and the last 6 h of the forecast. Individual mode probabilities were calculated based on how frequently a mode was forecast out of 48 possible occurrences for a 12-h period (four members $\times$ 12 h) and 24 possible occurrences for a 6-h periods (four members $\times$ 6 h). For example, if TS was forecast by only one member for the first 6 h of a simulation, the forecast probabilities would be 12.5% for the 12-h period (6/48), 25% for the first 6-h period (6/24), and 0% for the second 6-h period (0/24).

c. Forecast verification

The deterministic ensemble mode forecasts were verified using the accuracy score introduced in SG14. Time was normalized with the simulation initialization set to zero and the end of a simulation, or the dissipation of convection, set to one. The duration of each convective mode was represented as a fraction of this time. Hourly simulated convective modes were compared to the hourly observed convective modes and scored until the normalized time of 1.0 (minimum of 6 h and maximum of 12 h). When the simulated convective mode was a direct match with observations, a score of 1 was given. A score of 0.5 was given when the simulated convective mode matched the same group as observed but was not a direct match with the specific subtype. Finally, when there was no match, a score of 0 was given. The resulting accuracy scores ranged from 0 to 1 for each forecast, with 1.0 being a perfect forecast. The ensemble mode scores were then summed to determine the overall event accuracy score S using

$$S = \sum_{i=1}^{N} M\Delta t, \tag{1}$$

where N represents the total number of mode comparisons possible (a function of the number of times the mode changed in either the observations or the forecast), M is a weight based on the match type (direct, group, or no agreement) as described, and Δt is the normalized duration of the mode comparison (SG14). The forecasts were evaluated before and after correcting for the low bias in TS and BE forecasts.

The reliability of each convective mode was evaluated graphically for both the hourly direct and neighborhood forecast methods. Reliability is determined from a conditional distribution $[p(o|f)]$ indicating how often an observed convective mode o occurred when a particular forecast mode f was given. Forecasts have perfect reliability when

$$p(o = 1|f) = f, \tag{2}$$

(Murphy and Winkler 1987; Franz et al. 2003), that is, when the observed relative frequency of a convective mode equals the forecast probability for that mode (Murphy and Winkler 1992).

Because a 6- or 12-h time period can have multiple convective modes observed and forecast within the period, the divergence score (DS) and divergence skill score (DSS) introduced by Weijs et al. (2010) were also used to verify the probabilistic forecasts. The DS is calculated using the Kullback–Leibler (KL) divergence, which is a measure of relative entropy (Kullback and Leibler 1951). DS is similar to a Brier score (BS); however, the DS is asymmetric (whereas BS is symmetric) and incorporates multiple possible forecast solutions (whereas BS uses binary solutions) (Weijs et al. 2010). DS for a given forecast can be interpreted as the reduction of uncertainty in a probabilistic forecast from a climatological forecast (Weijs et al. 2010). DS is calculated as

$$\mathrm{DS} = D_{\mathrm{KL}}(o_i|f_i) = \sum_{i=1}^{n} o_i \ln\left(\frac{o_i}{f_i}\right), \tag{3}$$

which is read as divergence from f to o (KL divergence is directional) and is always positive for an entire forecast. In Eq. (3), n represents the total number of convective modes possible (nine), o represents the observed probability of convective mode i, and f represents the forecast probabilities of the same convective mode occurring over the 6- or 12-h periods, respectively. Climatological DS forecasts were calculated as

$$\mathrm{DS}_{\mathrm{ref}} = \frac{1}{N}\sum_{t=1}^{N} \mathrm{DS}, \tag{4}$$

where N is the total number of forecasts. Because of the limited sample size, an investigation into one or two DS values affecting the climatology too strongly was completed. Ten such potential outliers were found, but they did not significantly impact the results of our analysis.

A perfect forecast would result in a DS of zero with DS increasing as the uncertainty of a forecast increases. If DS of a particular forecast is equal to the climatological DS, then that forecast did not contain any more information than the climatology and essentially added no additional value (Weijs et al. 2010). A DS greater than climatology indicates an increase of uncertainty compared to climatology, and a DS between zero and the climatological value indicates a decrease of uncertainty compared to climatology in that particular forecast. It should be noted that a reduction (increase) in forecast uncertainty means that a forecast performed better (worse) than a climatological forecast. Similar to the BS, a procedure to normalize the DS relative to a climatological forecast was completed, yielding the DSS. Forecasts with numerous categories such as convective modes are calculated using (Weijs et al. 2010)

$$\mathrm{DSS} = 1 - \frac{\mathrm{DS}}{\mathrm{DS}_{\mathrm{ref}}}. \tag{5}$$

The DSS transforms a perfect forecast score from zero to one, and forecasts that are poorer than climatology become negative. DSSs were determined for each of the 12-h forecasts, the first 6 h of each forecast, and the last 6 h of the forecast.

Two issues arise with these equations when either the observed or forecast probability is zero. First, when a particular convective mode was forecast but not observed ($o_i = 0$) (Kullback and Leibler 1951), the result was treated as zero in the summation. This produces an artificial reduction of forecast uncertainty. Second, when a convective mode was observed but not forecast ($f_i = 0$) the equation became undefined. Therefore, a small epsilon value of 0.02 was introduced to eliminate these two issues (Han 2017). This value was chosen because a convective mode is only included when it lasts for at least two consecutive hours, and thus the lowest forecast probability for a convective mode is 0.042 for a 12-h forecast period [2/48 (12 h × four ensemble members)]. Although a convective mode is required to last for two consecutive hours, a convective mode could be predicted for hours 6 and 7 of a forecast period, resulting in the possibility of having a forecast convective mode occurring just once during a 6-h forecast period. In this case, the lowest forecast probability is also 0.042 [1/24 (6 h × four members)]. Furthermore, the epsilon value of 0.02 is small enough to induce a noticeable penalty for observed modes that were not forecast or forecast modes that did not occur as would be expected, but will prevent unreasonably large penalties from occurring.

The performance of individual convective mode forecasts was also evaluated using the DS and DSS. Although the DS for a given forecast including all convective modes is always positive, individual modes can have either positive or negative DS values. A positive (negative) DS mode value means the forecast probability of the mode is lower (higher) than the observed

climatological probability of the mode, indicating an underforecast (overforecast). Because KL divergence is not symmetric, it is difficult to determine what DS value represents a perfect forecast for a particular mode, thus, determining whether a DS value indicates an increased or decreased forecast uncertainty from climatology is also problematic. Therefore, the ratio of the absolute DS value of each forecast mode (DS_{mode}) to the absolute DS value for the study climatological forecast mode ($DS_{mode_{ref}}$) was used to calculate the DSS for each individual mode, or DSS_{mode},

$$DSS_{mode} = 1 - \frac{|DS_{mode}|}{|DS_{mode_{ref}}|}, \tag{6}$$

to determine whether uncertainty for that mode increased or decreased compared to climatology. A ratio of absolute DS values provides information about how much the overall uncertainty has increased or decreased, but fails to provide information about under- or overprediction of a particular convective mode.

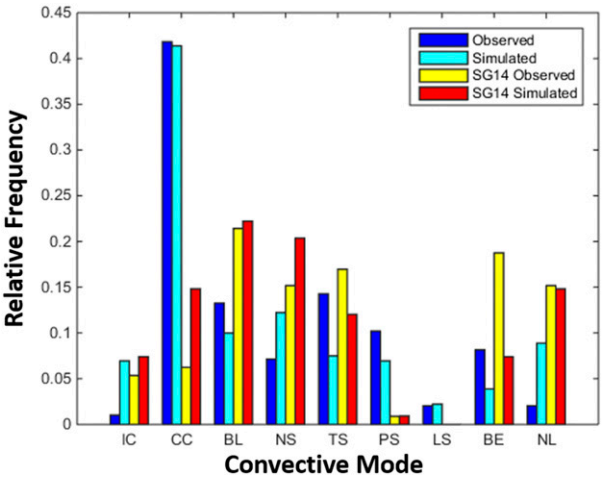

FIG. 3. Relative frequencies of observed and simulated convective modes (observed, blue; simulated, cyan) and SG14 (observed, yellow; simulated, red).

3. Results

Forecasts covered periods with 71 distinct observed convective modes and the four ensemble members simulated 360 different convective modes. However, the total number of simulated convective modes may not be thought of as a fully independent sample, as synoptic or mesoscale conditions can favor a particular mode, causing possible cross-member correlation in some forecasts. Forecasts produced higher relative frequencies than were observed for IC, NS, and NL, and lower relative frequencies than were observed for the remaining five convective modes (CC, BL, TS, PS, and BE) (Fig. 3). In addition, results for IC, NS, TS, and BE were similar to those in SG14 in which 115 observed convective modes and 109 simulated convective modes from the 2006 to 2010 warm seasons were examined. SG14 included three observed and one simulated mixed convective mode systems, but mixed-mode systems were not investigated in the present study and therefore were removed from the SG14 sample, resulting in 112 observed convective modes and 108 simulated convective modes for comparison. It should be noted that SG14 investigated systems occurring over the U.S. Great Plains and Upper Midwest rather than primarily over Iowa as in this study, and their simulations were run for 24 h rather than 12 h. Also, because they focused on deterministic runs from a single model configuration,

TABLE 3. Hourly ensemble modes of the seven cases comparing the small WRF and NCAR ensembles. A plus sign (+) indicates clear ensemble modes (>50% member agreement), and a minus sign (−) indicates ensemble modes requiring a tie breaker. Hours where the ensemble mode did not predict a convective mode are marked with an asterisk (*).

		1	2	3	4	5	6	7	8	9	10	11	12
4 May 2015	WRF	NS+	NS+	NS	TS	TS	CC+	CC+	CC+	CC+	CC+	IC−	IC
	NCAR	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC	BL	BL
5 May 2015	WRF	BE	BE+	BE+	BE	BE−	CC+	CC+	CC+	CC+	CC+		
	NCAR	CC+	CC	CC	CC+	CC	CC	CC	CC	BL−	BL−		
25 Jun 2015	WRF	CC+	CC+	CC+	CC−	BL	BL	BL−	CC−	CC	CC+	CC+	CC
	NCAR	CC+	CC	CC	BE−	BE	BL+	BL+	BL+	BL	CC+	CC+	CC+
29 Jul 2015	WRF	CC+	CC+	CC	BL+	BL+	CC+	CC+	CC+	*			
	NCAR	*	CC+	CC+	CC+	CC+	CC−	CC+	CC+	CC+			
10 Aug 2015	WRF	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+
	NCAR	*	CC	CC+	CC+	CC	CC	CC	CC+	CC+	CC+	CC+	CC+
17 Aug 2015	WRF	NS+	NS+	CC	CC−	CC	CC	CC+	CC+	CC+	CC+	CC+	CC+
	NCAR	CC	CC	CC	NS	NS−	NS	PS	PS−	CC	CC+	CC+	CC+
6 Apr 2016	WRF	CC−	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC+	CC	CC−	CC−
	NCAR	BL	BL	BL	CC+	CC+	CC+	CC+	CC+	CC+	IC+	IC+	IC

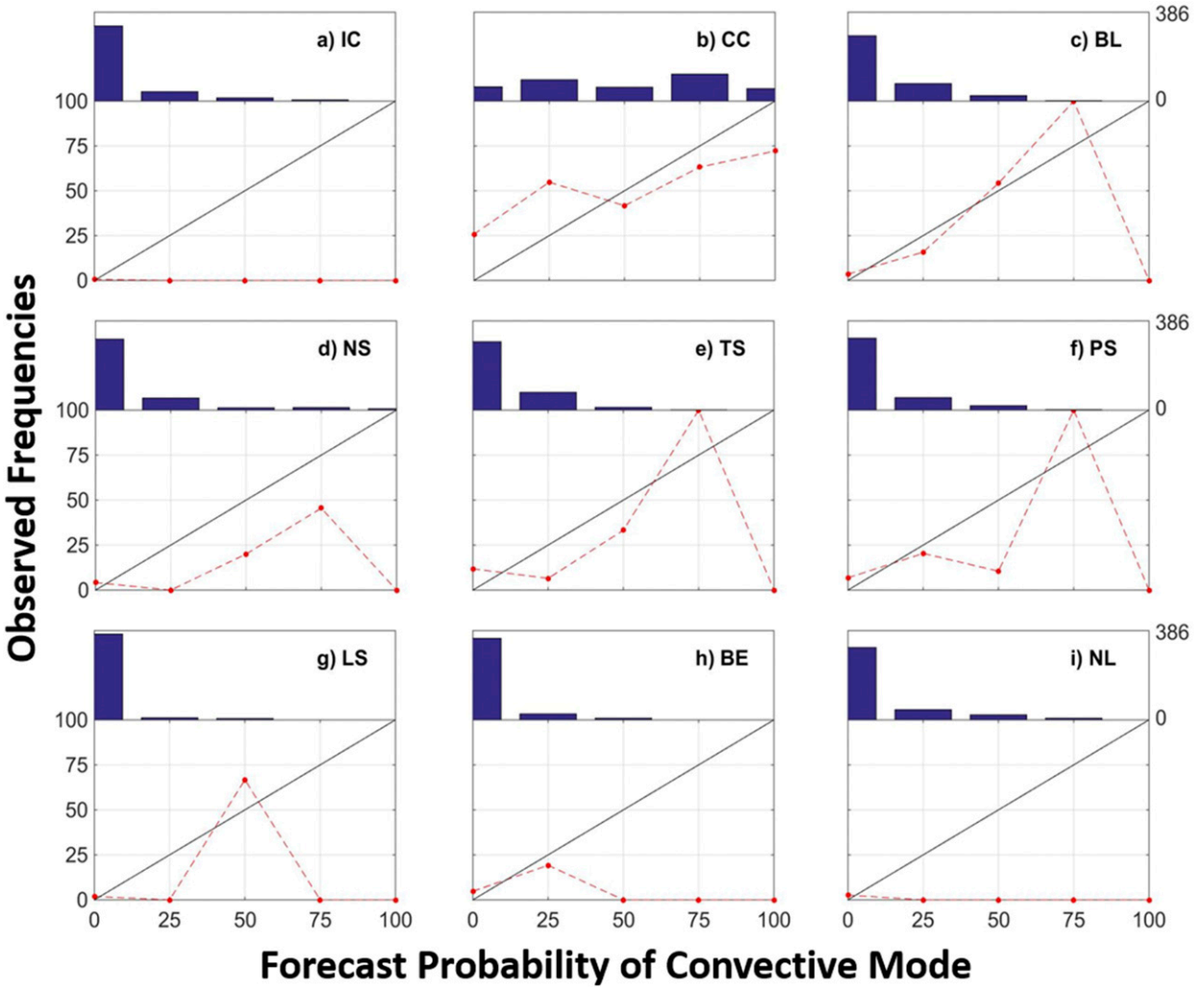

FIG. 4. Reliability diagrams for all nine convective modes using the direct forecasting method. Frequencies of occurrence for forecast probabilities are shown above each diagram.

there would naturally be some model configuration differences between the two studies (noted in Table 2).

TS and BE were the most underpredicted convective modes with TS events simulated 47.5% less frequently than observed and BE events simulated 52.4% less frequently than observed (Fig. 3). This result is similar to SG14. Done et al. (2004) also showed a tendency for convection-allowing WRF forecasts to underpredict TS systems (they did not classify BE events). CC events were observed and simulated much more frequently than the eight other convective modes (Fig. 3), with 27.5% more observed occurrences and 29.2% more simulated occurrences than the next most frequently occurring convective modes. In addition, the relative frequencies of simulated and observed CC events were similar with simulated events occurring only 1.1% less often than observed. This is in contrast to SG14, who reported that CC events were simulated 137% more frequently than observed.

The deterministic ensemble mode forecasts resulted in an average accuracy score of 0.50 with 60% of the forecasts scoring 0.50 or higher, both of which were higher than three of the four individual ensemble members. Despite not being a traditional ensemble mean, these results are consistent with findings that deterministic ensemble mean forecasts typically perform better than single, deterministic forecasts (Leith 1974; Fritsch et al. 2000; Baars and Mass 2005). For comparison, SG14 had an average score of 0.49 for 37 simulations with 41% of the forecasts scoring 0.50 or higher. Reclassifying the hourly ensemble mode convective modes to TS (BE) when at least two members

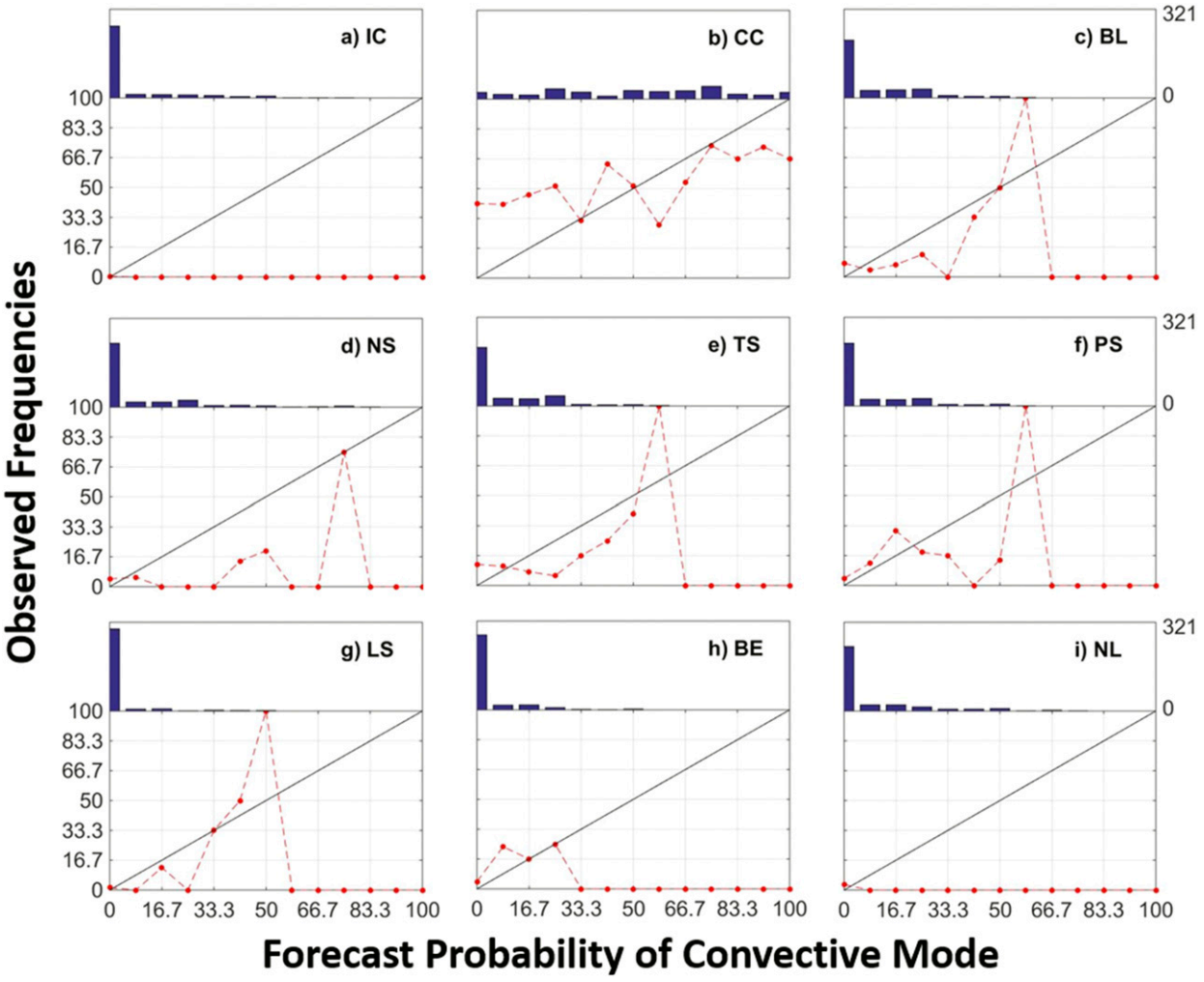

FIG. 5. As in Fig. 4, but using the neighborhood forecast approach.

were originally classified did not change the average accuracy score of 0.50 (only two forecasts were adjusted by this correction). The more liberal bias correction of reclassifying hourly ensemble mode forecasts to TS (BE) if at least one ensemble member was classified as TS (BE) resulted in an accuracy score of 0.41. This correction affected 20 forecasts, with 7 improved accuracy scores and 13 worsened accuracy scores.

For individual members, the NAM/NAM member resulted in a score of 0.57, while the three remaining members each scored 0.45. Similarly, the NAM/NAM member produced the highest percentage of forecasts scoring at least 0.50 at 66% compared to 43% for the GFS/GFS member, 37% for the NAM/GFS member, and 49% for the GFS/NAM member.

Despite it being much larger and having a more systematic approach in member creation, the skill of the NCAR ensemble was similar to the skill of the small WRF ensemble for predicting convective mode. The NCAR ensemble mode average score for the seven events investigated was 0.46 with 43% of the forecasts scoring at least 0.50, while the small WRF ensemble mode scored an average of 0.49 for the same events with 57% of the forecasts scoring at least 0.50. While the average mode from the smaller ensemble scored higher than 3 of the 4 members, the average NCAR ensemble mode scored higher than all 10 individual members.

The small WRF ensemble produced a clear convective mode (>50% member agreement) in 67% of the possible forecast hours and a tie among at least two members where the surrounding hours were used to determine the mode 10% of the time. Comparatively, the NCAR ensemble produced clear modes 62% of the time and ties 9%. In addition, the NCAR ensemble members produced a majority mode with less than 50% member agreement 29% of the time, which was slightly

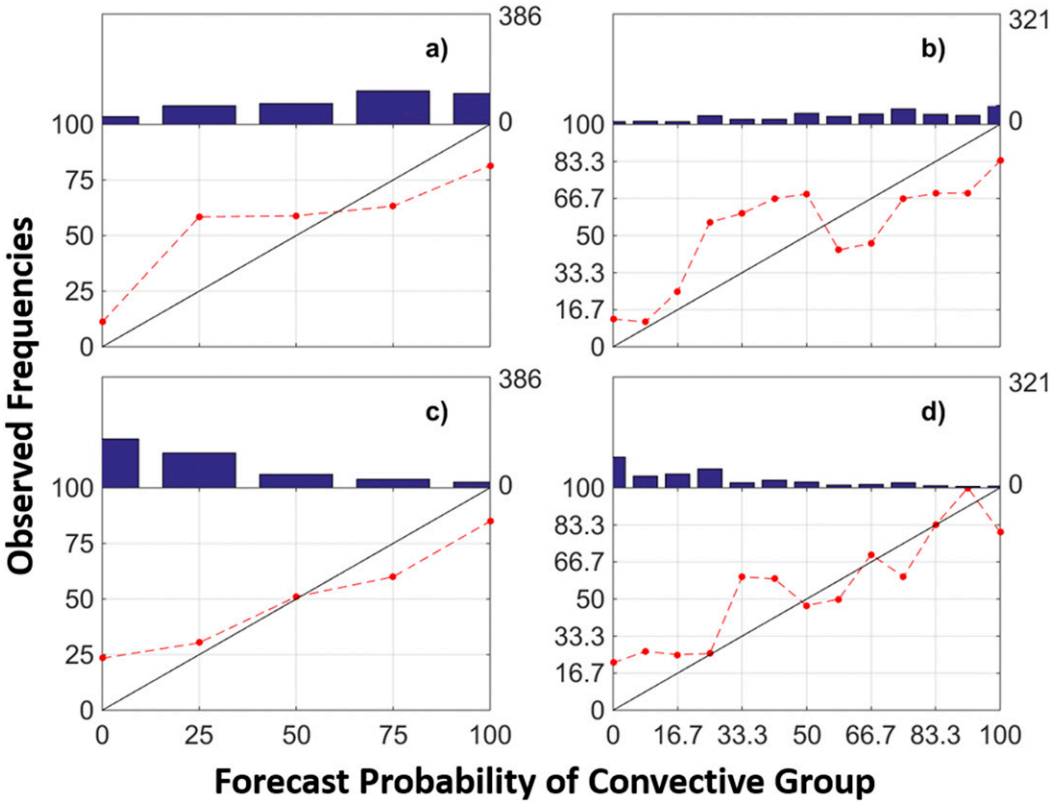

FIG. 6. Convective mode group reliability diagrams. Diagrams are as follows: (a) direct forecast method cellular, (b) neighborhood forecast method cellular, (c) direct forecast method linear, and (d) neighborhood forecast method linear.

more frequent than the WRF ensemble (23% of the time), suggesting that the larger ensemble may produce a wider spread of possible forecasts. Additionally, the same convective mode was forecast by both ensembles 54% of the time (Table 3). These results are encouraging and suggest that the simple, very small ensemble used in the present study may not behave substantially differently than larger ensembles typically being used for convection-allowing guidance by operational forecasters today. Although caution must be used as a result of the small sample size, the relative agreement suggests the results from the small, four-member WRF ensemble may be robust.

Reliability for the hourly forecast probabilities varied among the individual modes (Fig. 4). CC was the only convective mode that had a relatively large number of forecasts with 100% probability (54), and NS was the only other convective mode to have forecasts with 100% forecast probability (6). Forecasts of CC underpredicted for the 0% and 25% forecast probabilities and overpredicted for the 50%, 75%, and 100% forecast probabilities (Fig. 4b). BL forecasts were reliable up to the 50% forecast probability but were underpredicted at the 75% forecast probability. Forecasts of individual linear systems (NS, TS, PS, LS, and BE) had poor reliability with the exception of the 0% forecast probability (Figs. 4d–h). Although PS and BE forecasts were relatively reliable for 25% forecast probability (20.4% and 19.2% observed frequency, respectively), reliability declined for forecast probabilities greater than 25% for both modes. Forecasts for IC and NL were the least reliable (Figs. 4a,c). However, these two modes were the least frequently observed (2 h for IC and 8 h for NL) and, again, caution should be used when interpreting results because of the small sample size.

Reliability also varied among individual modes for the neighborhood forecast method (Fig. 5). Similar to the direct forecast approach, CC was the only convective mode to have forecasts that ranged across all forecast probability levels. Typically, CC was underforecast for probabilities up to 50% and overforecast for probabilities greater than 50% (Fig. 5b). With a few exceptions, BL, TS, and LS forecasts were fairly reliable up to the 50% forecast probability; however, forecast probabilities greater than 50% rarely or never occurred (Figs. 5c,e,g). Reliability of PS and BE declined with forecast probabilities

TABLE 4. Climatological DS and percentage of simulations where forecast uncertainty was reduced (increased skill) compared to a climatological forecast for the 0–12-, 0–6-, and 6–12-h forecast periods.

Forecast period (h)	Climatological DS	Percentage of forecasts with reduced uncertainty
0–12	0.84	68.8
0–6	1.28	65.6
6–12	1.02	66.7

greater than 25% (Figs. 5f,h), and the IC, NS, and NL forecasts were the least reliable (Figs. 5a,d,i).

Overall, the broader linear and cellular group forecasts were more reliable than those for individual convective modes for both the direct and neighborhood forecast methods (Fig. 6). Despite individual linear convective mode forecasts being less reliable than individual cellular mode forecasts, the linear convective group forecasts were more reliable than the cellular convective group forecasts, in contrast with the results found in SG14. These results suggest that a small ensemble can improve forecasts of linear convection; however, predicting the stratiform precipitation location remains difficult. In general for both convective groups, the observed frequency increased as the forecast probability increased. However, group convection was typically underestimated when forecast probabilities were less than 50%, and overestimated for forecast probabilities greater than 50% for both forecast methods, particularly for cellular convection (Fig. 6). With a few exceptions, the linear convection forecasts were fairly reliable, particularly with the neighborhood forecast method (Figs. 6c,d).

Of the three forecast periods investigated, the 12-h period had the lowest climatological DS (0.84), implying that the forecast convective modes for this period were more closely representative of the observed modes than the forecast modes for the 0–6- and 6–12-h periods (Table 4). Despite having the lowest climatological DS value, the 12-h forecast period had the highest percentage (68.8%) of forecasts with reduced uncertainty from a climatological forecast (Fig. 7), suggesting the ensemble may have more skill in simulating the relative frequency of a convective mode over a longer forecast period. The 0–6-h forecast period resulted in the lowest percentage of cases with reduced forecast uncertainty (65.6%), implying that the 0–6-h convective mode forecasts may not be an improvement over the 6–12-h forecasts; a result that seems in opposition to previous findings that radar data assimilation improves QPF skill and often most notably in the first 6–8 h of a forecast (Berenguer et al. 2012; Stratman et al. 2013). To explore

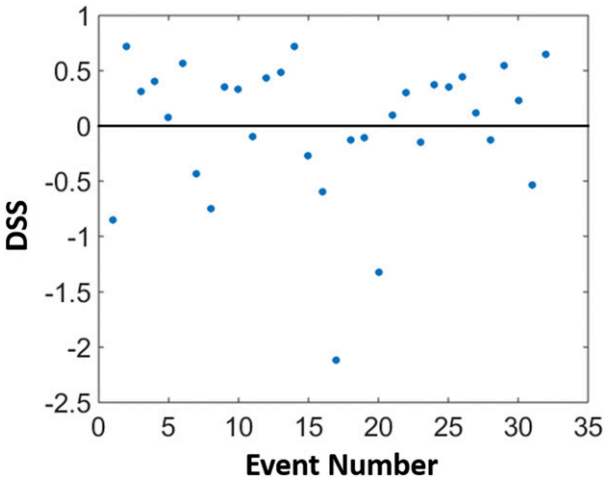

FIG. 7. The 12-h DSSs for all cases. DSSs greater than zero indicate a reduction in forecast uncertainty (increased skill) compared to climatology.

this issue further, simulations without radar data assimilation were run to examine the impact of the assimilation on convective mode forecasts during the first few hours (not shown). Simulations using radar data assimilation generally did forecast more vigorous convection during the first few hours than simulations without using assimilation, resulting in improved convective mode forecasts early in the period compared to when no assimilation was used. Thus, these results are similar to previous studies that investigated the effects of radar data assimilation on forecasting convection (e.g., Kain et al. 2010; Schenkman et al. 2011; Stratman et al. 2013; Moser et al. 2015). The fact that skill may not be as good in the first 6 h as it is for the 6–12-h forecast period is thus not the result of the assimilation harming the forecasts. Instead, it implies that the convective mode may be more predictable later in the lifetime of convective systems, even if QPF skill declines then. These results also suggest that although the radar data assimilation does reduce model spinup issues, the model might still require time to organize convection in a way that more closely matches the observations.

DSSs for individual convective modes were frequently equal to one, indicating that there were several perfect forecasts for each convective mode (Fig. 8; only the 12-h forecast period is shown). CC had the fewest DSS values equal to one while LS had the most DSS values equal to one, which is a function of how frequently the modes were forecast or observed. There were numerous instances when a given convective mode was neither forecast nor observed, producing several perfect forecasts of nonoccurring modes. This indicates that the ensemble generally performed well when forecasting

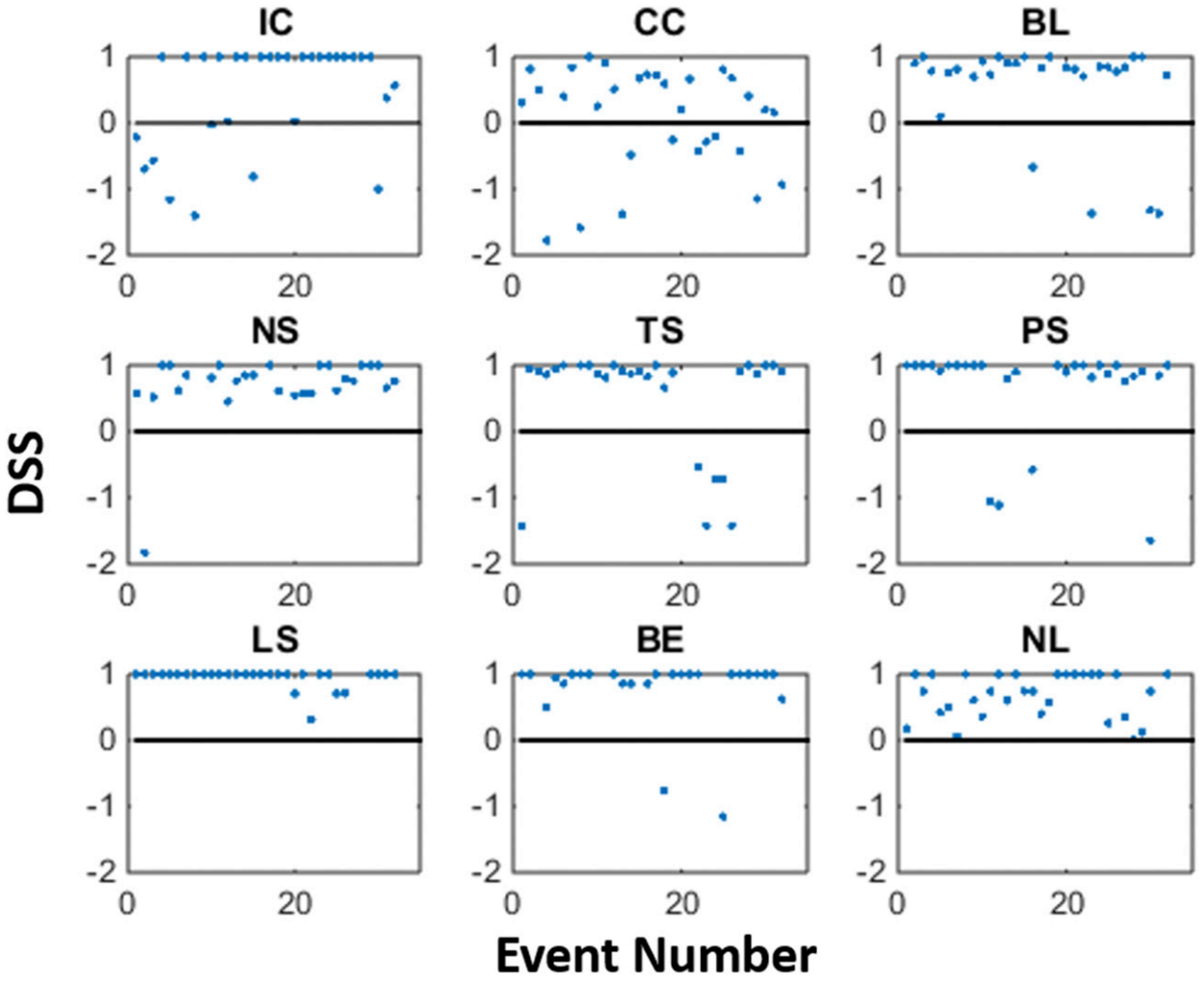

FIG. 8. The 12-h DSSs for all nine convective modes.

nonoccurring modes. Although it is important to understand how well nonoccurring modes are forecast, it is perhaps more important to assess how well the ensemble performs when a convective mode is either forecast or observed since these instances may be thought of as rare events.

When considering all three forecast periods (0–12, 0–6, and 6–12 h), the percentage of forecasts with reduced uncertainty from a climatological forecast was at least 56.3% for all convective modes (Table 5). In general, modes with better climatological forecasts (lower DSs) had higher percentages of forecasts with reduced uncertainty from their respective climatological forecasts, but this was not the case for all situations. When the nonobserved, nonforecast modes were excluded, the percentage of forecasts with reduced uncertainty from a climatological forecast decreased for every convective mode except CC (Table 6), suggesting the skill of the ensemble declines when forecasted modes do occur. For a better overall comparison, the same climatological DS values were used for this subset of events. IC averaged the largest decrease in percentage (70.1%) of forecasts with reduced uncertainty. Because CC was either observed or forecast for every event, there was no change in the forecast uncertainty. Note that the LS forecast probabilities exactly matched the observed probabilities for all of the 6–12-h forecast periods; however, only one of the 32 simulations had forecast and observed an LS during this period. Therefore, the LS forecast sample size for the 6–12-h period was too small to draw meaningful conclusions (Tables 5 and 6).

4. Conclusions

One deterministic ensemble mode forecast and multiple probabilistic forecasts were investigated using a

TABLE 5. As in Table 4, but for individual convective modes. A plus sign (+) [minus sign (−)] indicates the convective mode producing the highest (lowest) percentage of improved forecasts compared to a climatological forecast for each forecast period. The asterisk (*) denotes a sample size too small to draw meaningful conclusions for the LS 6–12-h forecast period.

	Climatological DS			Percentage of forecasts with reduced uncertainty		
Convective mode	0–12 h	0–6 h	6–12 h	0–12 h	0–6 h	6–12 h
IC	0.02	0.03	0.02	71.8	90.6	63.6−
CC	0.23	0.31	0.38	62.5−	75.0	66.7
BL	0.15	0.19	0.16	78.1	78.1	84.8
NS	0.10	0.12	0.09	84.4	93.8+	84.8
TS	0.26	0.34	0.28	71.8	81.3	84.8
PS	0.22	0.27	0.22	78.1	84.4	84.8
LS	0.05	0.12	0	93.8	93.8+	*
BE	0.10	0.18	0.07	75.0	81.3	93.9+
NL	0.05	0.04	0.11	96.9+	56.3-	93.9+

sample of thirty-two 12-h model simulations to determine the ability of a small WRF ensemble to predict short-term convective mode evolution. Each simulation was initialized with assimilated radar data and used 3- or 6-h forecast output from the NAM and GFS models for initial and lateral boundary conditions to allow for use in a real-time situation. Convection from these events was classified into one of nine modes—three cellular, five linear, and one nonlinear—as described in G08. Overall, the ensemble produced an overestimation of IC, NS, LS, and NL systems, while underestimating CC, BL, TS, PS, and BE systems.

Deterministic forecasts were created using the statistical mode of the ensemble for each hour. Because the ensemble is smaller than what would likely be used operationally, the 10-member NCAR ensemble was used as comparison for seven events for which data were available from both ensembles. In addition, two simple bias corrections were tested to account for the tendency of TS and BE to be underpredicted for the small ensemble (Done et al. 2004; SG14). These deterministic forecasts were verified by using the accuracy score introduced by SG14. Probabilistic forecasts were created for 1-, 6-, and 12-h time periods. Hourly probabilistic forecasts of individual convective modes and the broader convective groups using a direct forecast method (mode probabilities calculated using only that particular forecast hour) and a neighborhood forecast method (mode probabilities calculated also using the hour prior and the hour after the time of interest) were verified using reliability. The 6- and 12-h forecast probabilities were verified using divergence scores and divergence skill scores, measures of forecast uncertainty compared to a climatological forecast (Weijs et al. 2010).

Despite the small size of the WRF ensemble used in the present study, it does show skill in forecasting convective mode evolution. Analysis of the larger NCAR ensemble revealed similar skill, indicating that a small ensemble with diverse members could be as useful in forecasting the convective mode as a larger ensemble. These results also suggest the analysis of the small WRF ensemble may be generalizable to other ensemble systems. The approach using the statistical

TABLE 6. As in Table 5, but only including the events with an observed or simulated convective mode.

	Climatological DS			Percentage of simulations with information gained		
Convective mode	0–12 h	0–6 h	6–12 h	0–12 h	0–6 h	6–12 h
IC	0.02	0.03	0.02	12.5−	3.1−	0−
CC	0.23	0.31	0.38	62.5+	75.0+	66.7+
BL	0.15	0.19	0.16	59.4	34.4	45.5
NS	0.10	0.12	0.09	59.4	43.8	27.3
TS	0.26	0.34	0.28	46.9	37.5	39.4
PS	0.22	0.27	0.22	40.6	25	33.3
LS	0.05	0.12	0	21.9	12.5	*
BE	0.10	0.18	0.07	28.1	9.4	15.2
NL	0.05	0.04	0.11	59.4	6.3	18.2

mode for deterministic ensemble forecasts also exhibited a general improvement in skill over the individual members with both ensembles. Similar to previous studies (Done et al. 2004; SG14), TS and BE forecasts were among the least accurate. However, simple bias corrections accounting for the underprediction of TS and BE did not improve the forecast skill, and future work should explore more sophisticated methods that might take into account near-storm environmental conditions.

Probabilistic forecasts of individual convective modes showed some reliability, with broader cellular and linear convective groups having more reliable forecasts than individual convective modes. However, a tendency for the ensemble to underpredict convective groups for probabilities less than 50% and overpredict convective groups for probabilities greater than 50% suggests forecast overconfidence. Consistent with the behavior of precipitation forecasts, the results indicate that the ensemble has better skill forecasting convective modes over the longer 12-h period than either 6-h period.

Given that the present study is preliminary and focuses on the Iowa vicinity, it is unknown how well the ensemble will predict convective modes in different regions, such as a mountainous area. Future work should explore the performance of the ensemble in predicting convective modes in other regions where the land characteristics might be more heterogeneous. It is also unclear how much sensitivity exists between the correlation among ensemble members and the particular synoptic conditions. Thus, future work should investigate the role of synoptic setup on the convective mode forecasts. Future work should also explore additional ways to produce deterministic convective mode forecasts from an ensemble.

Acknowledgments. The authors thank Dr. Keith Brewster at the Center for Analysis and Prediction of Storms and Tim Supinie at the University of Oklahoma for their assistance with the ARPS program and Dr. Mark Kaiser at Iowa State University for suggesting the use of KL-Divergence as a statistical evaluation tool. Daryl Herzmann and Dave Flory assisted with the computational resources. This work was supported by NOAA CSTAR Awards NA14NWS4680016 and NA17NWS4680005, and NSF Award AGS1624947. This project also took advantage of netCDF software developed by UCAR/Unidata (http://doi.org/10.5065/D6H70CW6) and the experimental ensemble produced by NCAR (https://ensemble.ucar.edu/). The authors appreciate the constructive comments from three anonymous reviewers.

REFERENCES

Baars, J. A., and C. F. Mass, 2005: Performance of National Weather Service forecasts compared to operational, consensus, and weighted model output statistics. *Wea. Forecasting*, **20**, 1034–1047, https://doi.org/10.1175/WAF896.1.

Baldwin, M. E., J. S. Kain, and S. Lakshmivarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.*, **133**, 844–862, https://doi.org/10.1175/MWR2892.1.

Berenguer, M., M. Xue, and F. Kong, 2012: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison among numerical models and with nowcasting. *Mon. Wea. Rev.*, **140**, 2689–2705, https://doi.org/10.1175/MWR-D-11-00181.1.

Bluestein, H. B., and M. H. Jain, 1985: Formation of mesoscale lines of precipitation: Severe squall lines in Oklahoma during the spring. *J. Atmos. Sci.*, **42**, 1711–1732, https://doi.org/10.1175/1520-0469(1985)042<1711:FOMLOP>2.0.CO;2.

Brewster, K., 1996: Application of a Bratseth analysis scheme including Doppler radar data. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 92–95.

CAPS, 2017: Model initialization and preprocessing. ARPS User's Guide, Version 4.0, Center for Analysis and Prediction of Storms, 214–240, http://www.caps.ou.edu/ARPS/download/code/pub/ARPS.docs/ARPS4DOC.PDF/arpsch8.pdf.

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, https://doi.org/10.1175/2009WAF2222222.1.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, https://doi.org/10.1002/asl.72.

Doswell, C. A., III, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581, https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.

Duda, J. D., and W. A. Gallus Jr., 2010: Spring and summer midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206, https://doi.org/10.1175/2009WAF2222338.1.

——, and ——, 2013: The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in the WRF. *Wea. Forecasting*, **28**, 994–1018, https://doi.org/10.1175/WAF-D-13-00005.1.

Elmore, K. L., S. J. Weiss, and P. C. Banacos, 2003: Operational ensemble cloud model forecasts: Some preliminary results. *Wea. Forecasting*, **18**, 953–964, https://doi.org/10.1175/1520-0434(2003)018<0953:OECMFS>2.0.CO;2.

Fowle, M., and P. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794, https://doi.org/10.1175/1520-0434(2003)018<0782:SHNPOC>2.0.CO;2.

Franz, K. J., H. C. Hartmann, S. Sorooshian, and R. Bales, 2003: Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin. *J. Hydrometeor.*, **4**, 1105–1118, https://doi.org/10.1175/1525-7541(2003)004<1105:VONWSE>2.0.CO;2.

Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky, 2000: Model consensus. *Wea. Forecasting*, **15**, 571–582, https://doi.org/10.1175/1520-0434(2000)015<0571:MC>2.0.CO;2.

Fujita, T. T., 1978: Manual of downburst identification for Project NIMROD. SMRP Research Paper 156, University of Chicago, 104 pp. [NTIS N78-30771/7GI.]

Gallus, W. A., Jr., N. A. Snook, and E. V. Johnson, 2008: Spring and summer severe weather reports over the Midwest as a function of convective mode: A preliminary study. *Wea. Forecasting*, **23**, 101–113, https://doi.org/10.1175/2007WAF2006120.1.

Grams, J. S., W. A. Gallus Jr., S. E. Koch, L. S. Wharton, A. Loughe, and E. Ebert, 2006: The use of a modified Ebert–McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. *Wea. Forecasting*, **21**, 288–306, https://doi.org/10.1175/WAF918.1.

Han, J., 2017: Kullback–Leibler divergence. Accessed 15 May 2017, http://web.engr.illinois.edu/~hanj/cs412/bk3/KL-divergence.pdf.

Hilgendorf, E. R., and R. H. Johnson, 1998: A study of the evolution of mesoscale convective systems using WSR-88D data. *Wea. Forecasting*, **13**, 437–452, https://doi.org/10.1175/1520-0434(1998)013<0437:ASOTEO>2.0.CO;2.

Homar, V., D. J. Stensrud, J. J. Levit, and D. R. Bright, 2006: Value of human-generated perturbations in short-range ensemble forecasts of severe weather. *Wea. Forecasting*, **21**, 347–363, https://doi.org/10.1175/WAF920.1.

Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.

——, Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Hu, M., M. Xue, J. Gao, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 699–721, https://doi.org/10.1175/MWR3093.1.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF Model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181, https://doi.org/10.1175/WAF906.1.

——, and Coauthors, 2010: Assessing advances in the assimilation of radar data and other mesoscale observations within a collaborative forecasting–research environment. *Wea. Forecasting*, **25**, 1510–1521, https://doi.org/10.1175/2010WAF2222405.1.

Kullback, S., and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86, https://doi.org/10.1214/aoms/1177729694.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418, https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014.

Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, https://doi.org/10.1175/2008MWR2556.1.

Moser, B. A., W. A. Gallus Jr., and R. Mantilla, 2015: An initial assessment of radar data assimilation on warm season rainfall forecasts for use in hydrologic models. *Wea. Forecasting*, **30**, 1491–1520, https://doi.org/10.1175/WAF-D-14-00125.1.

Murphy, A. H., 1991: Probabilities, odds, and forecasts of rare events. *Wea. Forecasting*, **6**, 302–307, https://doi.org/10.1175/1520-0434(1991)006<0302:POAFOR>2.0.CO;2.

——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338, https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2.

——, and ——, 1992: Diagnostic verification of probability forecasts. *Int. J. Forecasting*, **7**, 435–455, https://doi.org/10.1016/0169-2070(92)90028-8.

Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, https://doi.org/10.2151/jmsj.87.895.

Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 3413–3436, https://doi.org/10.1175/1520-0493(2001)129<3413:OMOMMC>2.0.CO;2.

Pettet, C. R., and R. H. Johnson, 2003: Airflow and precipitation structure of two leading stratiform mesoscale convective systems determined from operational datasets. *Wea. Forecasting*, **18**, 685–699, https://doi.org/10.1175/1520-0434(2003)018<0685:AAPSOT>2.0.CO;2.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011: The analysis and prediction of the 8–9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139**, 224–246, https://doi.org/10.1175/2010MWR3336.1.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, https://doi.org/10.1175/WAF-D-15-0103.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., http://dx.doi.org/10.5065/D68S4MVH.

Smith, B. T., R. Edwards, A. R. Dean, and R. L. Thompson, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part III: Tropical cyclone tornadoes. *Wea. Forecasting*, **27**, 1507–1519, https://doi.org/10.1175/WAF-D-11-00117.1.

Snively, D. V., and W. A. Gallus Jr., 2014: Prediction of convective morphology in near-cloud-permitting WRF Model simulations. *Wea. Forecasting*, **29**, 130–149, https://doi.org/10.1175/WAF-D-13-00047.1.

Stratman, D. R., M. C. Coniglio, S. E. Koch, and M. Xue, 2013: Use of multiple verification methods to evaluate forecasts of convection from hot- and cold-start convection-allowing models. *Wea. Forecasting*, **28**, 119–138, https://doi.org/10.1175/WAF-D-12-00022.1.

Sukoriansky, S., B. Galperin, and I. Staroselsky, 2005: A quasi-normal scale elimination model of turbulent flows with stable stratification. *Phys. Fluids*, **17**, https://doi.org/10.1063/1.2009010.

Tao, W.-K., J. Simpson, and M. McCumber, 1989: An ice-water saturation adjustment. *Mon. Wea. Rev.*, **117**, 231–235, https://doi.org/10.1175/1520-0493(1989)117<0231:AIWSA>2.0.CO;2.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

Wandishin, M. S., M. E. Baldwin, and S. L. Mullen, 2005: Short-range ensemble forecasts of precipitation type. *Wea. Forecasting*, **20**, 609–626, https://doi.org/10.1175/WAF871.1.

Weijs, S. V., R. Van Nooijen, and N. Van de Giesen, 2010: Kullback–Leibler divergence as a forecast skill score with classic reliability–resolution–uncertainty decomposition. *Mon. Wea. Rev.*, **138**, 3387–3399, https://doi.org/10.1175/2010MWR3229.1.

Yan, H., and W. A. Gallus Jr., 2016: An evaluation of QPF from the WRF, NAM, and GFS models using multiple verification methods over a small domain. *Wea. Forecasting*, **31**, 1363–1379, https://doi.org/10.1175/WAF-D-16-0020.1.